

## SUMMARY OF 1971 PATTERN RECOGNITION PROGRAM DEVELOPMENT

by

Sidney Lanier Whitley  
National Aeronautics and Space Administration  
Earth Resources Laboratory  
Mississippi Test Facility

INTRODUCTION

This paper is a summary of activities in the area of digital Pattern Recognition Analysis Programs at the Earth Resources Laboratory at the Mississippi Test Facility during 1971. This paper addresses eight areas related to Pattern Recognition Analysis at the Earth Resources Laboratory as follows: (1) Background; (2) Earth Resources Laboratory Goals; (3) Software Problems/Limitations; (4) Operational Problems/Limitations; (5) Immediate Future Capabilities; (6) Earth Resources Laboratory Data Analysis System; (7) General Program Needs and Recommendations; (8) Schedule and milestones.

BACKGROUND

In the mid-sixties when the Earth Resources Survey Program was much smaller, the University of Michigan, Purdue University, and the University of Kansas developed basically similar techniques for the analysis of multispectral data. The techniques developed by Michigan and Purdue were oriented toward multispectral scanner data analysis while the Kansas technique was oriented toward all kinds of imagery analysis; i.e., photographic, radar imagery, infrared scanner imagery, etc. The Pattern Recognition Analysis Technique implemented by the University of Michigan was, in the early days, an analog computer solution. Its chief limitations were long setup time, and dependence upon operator judgement. Its desirable feature was that, after long setup time, the classification run was very rapid.

The Pattern Recognition Analysis Technique developed at Purdue University was a series of algorithms which were implemented on an IBM 360/44 Computer. Figure 1 is a flow diagram describing the general flow of multispectral scanner data through the Purdue University Laboratory for Applications of Remote Sensing (LARS) Pattern Recognition Analysis System.

The analog tape-recorded multispectral scanner data were digitized and fed through a series of programs whose functions are as follows:

- PICTOUT - To display an image on a conventional line printer to simulate gray shades with selected symbols. This was used to identify training samples, test fields, and areas to be analyzed in detail.
- STAT - To compute and print on a line printer the following parameters for training classes and fields only:
- o Means
  - o Covariance matrices
  - o Histograms
  - o Spectral plots
  - o Coincident spectral plots
- SELECT - To determine the best  $n$  bands where  $2 < n < 12$  by a divergence calculation.
- CLASSIFY - To classify unknown materials by computing the probability that the unknown material is the same as one of the training samples for which a means and covariance has been stored as a signature.
- DISPLAY - To print the results from classify as a coded computer generated map, and to generate a score card to tell the investigator how well his run worked in designated test fields.

The Purdue Pattern Recognition Programs were comparatively easy to set up, were very repeatable when identical inputs were used and equipment operator judgement was at a minimum.

In late 1969, MSC/Houston acquired the Purdue Pattern Recognition Programs, implemented them on an IBM 360/44 Computer at the Manned Spacecraft Center, Houston, and developed a set of documentation for the series of programs. To demonstrate the complexity of the programs, it took approximately 9 months to get the system operational on a very similar computer. Some of the problems were:

- a. Systems routines delivered by the computer manufacturer contained errors. These particular systems routines had never been used by MSC programmers before, and the bugs had not been discovered.
- b. A small number of coding errors were found in the programs.
- c. Test case results did not match the results presented in Purdue's reports. It was finally determined that the report results were slightly in error.

Dr. David Landgrebe, Director, Laboratory for Applications of Remote Sensing at Purdue University, once stated that it was not sufficient to just have a correctly operating Digital Pattern Recognition Program, but that a lot of knowledge was required to run it effectively. That is a very true statement. To effectively operate the system, a team of engineers, computer programmers, statisticians and discipline scientists are required. An expertise in operating the programs comes with experience and careful study.

## EARTH RESOURCES LABORATORY GOALS

Figure 2 is a list of the objectives of the Earth Resources Laboratory. As discussed in a previous paper, a major interest of the Earth Resources Laboratory is to apply current data analysis techniques to experimentally demonstrate their applicability in the test area around the Mississippi Test Facility. Our interest was not to develop new programs and analysis techniques, but to use and evaluate previously established techniques, and to modify them as necessary to meet our needs.

To meet these objectives of the Earth Resources Laboratory we acquired copies of the latest versions of the Manned Spacecraft Center's Pattern Recognition Programs in May and June of 1971. By that time, MSC had begun the process of converting the Pattern Recognition Programs to run on UNIVAC 1108 Computers. At the time ERL acquired the UNIVAC 1108 programs, they still contained a number of errors caused by incompatibility between the IBM 360/44 and the UNIVAC 1108.

As ERL personnel started to convert the Digital Pattern Recognition Programs to run on the UNIVAC 1108 Computer at the Slidell Computing Center, we felt that we needed to familiarize our analysis personnel with the operational procedures, capabilities, and limitation of the programs. To accomplish this, we established a Pilot Automated Land Use Experiment following the step-by-step procedure shown in Figure 3. The area selected for a test case is shown in Figure 4, and is located in Harrison County, Mississippi. We used a frame of Color IR film for our source of data. We digitized our data with a scanning microdensitometer making tri-color separations to give us our multispectral data. We exercised all of our pattern recognition sub-programs with this pilot set of data. Figure 5 shows the flow of the data through the initial programs. Figure 6 gives an example of the PICTOUT display of one of the channels. Figures 7, 8, and 9 are examples of part of the output from Program STAT. Figure 10 is an example of the output from Program CLASSIFY, and Figure 11 is an example of output from program DISPLAY. Figure 12 is a score card showing the results of the classification.

Our first version of the program worked well with three channels of data. We moved on at the completion of the pilot study and software conversion to complete preparation for handling our first set of data from the 24-band Multispectral Scanner. Mr. Mooneyhan has shown a sample result of the first run of Multispectral Scanner data through ERL's pattern recognition programs.

#### SOFTWARE PROBLEMS/LIMITATIONS

In our preparation for handling 24 channel (MSDS) Multispectral Scanner Data, we found that there were many problems in the programs. The programs we received from MSC had a capability of reading MSDS data with 24 channels, and 700 elements per scan line, but a subset of 12 channels and 222 picture elements were all that the programs could process. At this point ERL assessed the tasks of modification of the programs to handle MSDS data before a C-130 MSDS checkout mission was flown. A decision was made to use an interim procedure for processing MSDS checkout mission data where the 700 element scan lines were divided into three parts; left third, middle third, and right third. After completion of the data processing, these segments are mosaiced together to reconstruct the entire swath width. Appendix A describes the interim data flow, assumptions, investigator participation requirements, computer processes and computer time estimate for using the interim procedure for processing 30,000 scan lines of data.

A major limitation of the digital pattern recognition programs is long computer run time due to its many calculations. Although there are a number of equations used in the series of programs, the following listed equation seems to be the greatest user of computer time. The following equation is used to compute the probability that material  $i$  gave rise to the measurement vector  $X$  for a particular ground cell or picture element in the multispectral imagery:

$$P_i(X) = \frac{1}{(2\pi)^{N/2} |K_i|^{1/2}} \exp \left[ -\frac{1}{2} (X-M_i)^T K_i^{-1} (X-M_i) \right]$$

Where:  $P_i$  = Probability density function from training fields for the  $i^{\text{th}}$  material  
 $X$  = Measurement vector for a particular ground cell  
 $N$  = Number of channels used  
 $M_i$  = Mean vector for the  $i^{\text{th}}$  material  
 $K_i$  = Covariance matrix for the  $i^{\text{th}}$  material

$K_i$  is an  $N \times N$  matrix where  $N$  is the number of channels used in the solution.  $P_i(X)$  is evaluated for each material for which a training sample is designated. This evaluation causes the matrix  $K_i$  to be

manipulated for every picture element in the scene for every material to be classified.

#### OPERATIONAL PROBLEMS/LIMITATIONS

In addition to the software problems/limitations described above, we have found some operational problems and limitations which are listed below:

- a. All Multispectral Scanner Data from the 24-band Scanner, and the S-192 EREP Scanner must be funneled through a Data Analysis System prior to computer analysis.
- b. Data Analysis Systems produce only 9-track computer tapes, causing users with only 7-track tape drives to have to procure 9-track tape drives or have the tapes converted.
- c. Software, Hardware, and procedural bugs still exist in the 24-channel Scanner and its Data Analysis System. This is not intended as a criticism, as the system is still in checkout. We have been able to work around all problems encountered to date.
- d. Interim display capabilities being used by ERL in the MTF area for computer generated color coded maps are cumbersome to use.

#### IMMEDIATE FUTURE CAPABILITIES

In view of the software problems/limitations discussed above, we have begun to modify and extend our series of pattern recognition programs, as necessary, to handle data from the newer multispectral scanners. Program STAT has been modified to accept Multispectral Scanner training sample data which is formatted by a Data Analysis System. STAT can now handle up to 24 bands of data and up to 700 picture elements per scan line. Program SELECT is being modified to accept up to 24 bands of Multispectral Scanner Data. We intend to always use fewer than 12 bands of data in our CLASSIFY program due to the long run time. We are still evaluating whether we should modify program CLASSIFY to accept 700 element per scan line data or if we should use our much faster Digital Table Look-Up classification, which gives equivalent results.

In addition to these software problems, digital pattern recognition programs have a basic problem of running slowly. A little over

a year ago an effort was initiated at MSC to develop a classification technique which would perform a pattern recognition analysis at a much faster rate and with results comparable to the Purdue Programs. Dr. Walter G. Eppler, while an LEC Direct Support Contractor at MSC, conceived a new Digital Table Look-Up pattern recognition technique which was many times faster than the LARS Pattern Recognition Analysis Programs. The Digital Table Look-Up (DTL) avoided computations by looking up the identity of materials from stored decision tables. These stored decision tables were based upon the Maximum Likelihood decision rule used in the Purdue Pattern Recognition programs. Dr. Eppler reported on the Table Look-Up Approach to Pattern Recognition at the Seventh International Symposium on Remote Sensing of Environment in May 1971. The DTL was implemented on an IBM 360/44 Computer at MSC as a machine language program.

A Digital Table Look-Up Pattern Recognition Analysis Program is being developed by the Earth Resources Laboratory at MTF for use on a UNIVAC 1108 Computer. Figure 13 shows the flow of data through the Digital Table Look-Up Program. It is estimated that this series of programs will be checked out and ready for operational verification by July 1972.

#### EARTH RESOURCES LABORATORY DATA ANALYSIS SYSTEM

The important role played by the Data Analysis Systems in preparing Multispectral Scanner Data for computer processing by users has been pointed out already. The format of flight tapes are not compatible with conventional data processing facilities. The Earth Resources Laboratory is acquiring a Data Analysis System to be installed at MTF because we are planning to emphasize the use and applications of Multispectral Scanner data acquired from aircraft and space. The scheduled delivery date for our DAS is in May 1972, with an operational date of about one month later. Figure 14 is an artists conception of the Data Analysis System. The ERL-DAS is similar to the two existing DAS's at MSC. It is completely compatible with the two MSC DAS's and has a number of features that do not exist in the MSC versions. These features, for the greatest part, were added at little or no cost, and were based upon experience gained in the design, manufacture, and checkout of the MSC-DAS's which were earlier models. The ERL DAS has one major capability that does not exist in the MSC DAS's. The ERL DAS can read, register, and display on the DAS up to four bands of Multiband Photography and process the data as though it were Multispectral Scanner Data. The same optical input subsystem can separate the three dye layers on color or color IR film and input them into the DAS as three-band multispectral data.

During the development and checkout of the two MSC-DAS's it was determined that the cursor system which was designed to designate training samples, test fields, and areas to be analyzed was more cumbersome to operate than necessary. To make the cursor system more convenient to use, the ERL DAS employs a light pen cursor.

Figure 15 is a functional block diagram of the ERL-DAS showing inputs, data processor, control and display, and output subsystems. Figure 16 gives the ERL-DAS Characteristics.

The ERL-DAS will be delivered with the necessary software to screen, evaluate, and reformat data from the 24-band Multispectral Scanner, the S-192 Skylab EREP Scanner, S-190 Multiband Photography (or any multiband photography), and Microwave Imager Data. The necessary software will be included to convert the S-192 curved scan lines to straight scan lines.

#### GENERAL PROGRAM NEEDS AND RECOMMENDATIONS

The Digital Pattern Recognition Programs are, as pointed out earlier, designed for processing 222 picture elements per scan line and twelve or fewer data bands. It should be recalled that the following numbers of scan lines must be processed in the near future:

<u>SCANNER</u>	<u>ELEMENTS/SCAN</u>	<u>DATA BANDS</u>	<u>SCAN RATE</u>
MSDS	700	24	10 to 100 scans/sec.
ERTS Scanner	2400	4	
SKYLAB EREP S-192	1200	12	100 scans/sec.

Prior to the time of the ERTS flight we have a need to modify our Pattern Recognition Programs to give them the capability to handle as many as 2400 picture elements per scan line.

Due to the large amount of resources required to analyze multispectral data it is highly desirable to make some preliminary checks to determine the quality of the data before proceeding with the full analysis. Since we have so much data these checks must be made automatically. For instance, tests can be run on each training sample to determine if it is sufficiently like other training samples of the same material, and if it is significantly different, the training sample can be rejected on the basis of investigator's criteria. If the data distribution is significantly different from a normal distribution, which is assumed by the maximum likelihood ratio classifier, the investigator can be given an opportunity to determine if processing should continue. At this point preprocessing algorithms such as

the ones previously used by the University of Michigan could be applied so that the analysis results will be more correct.

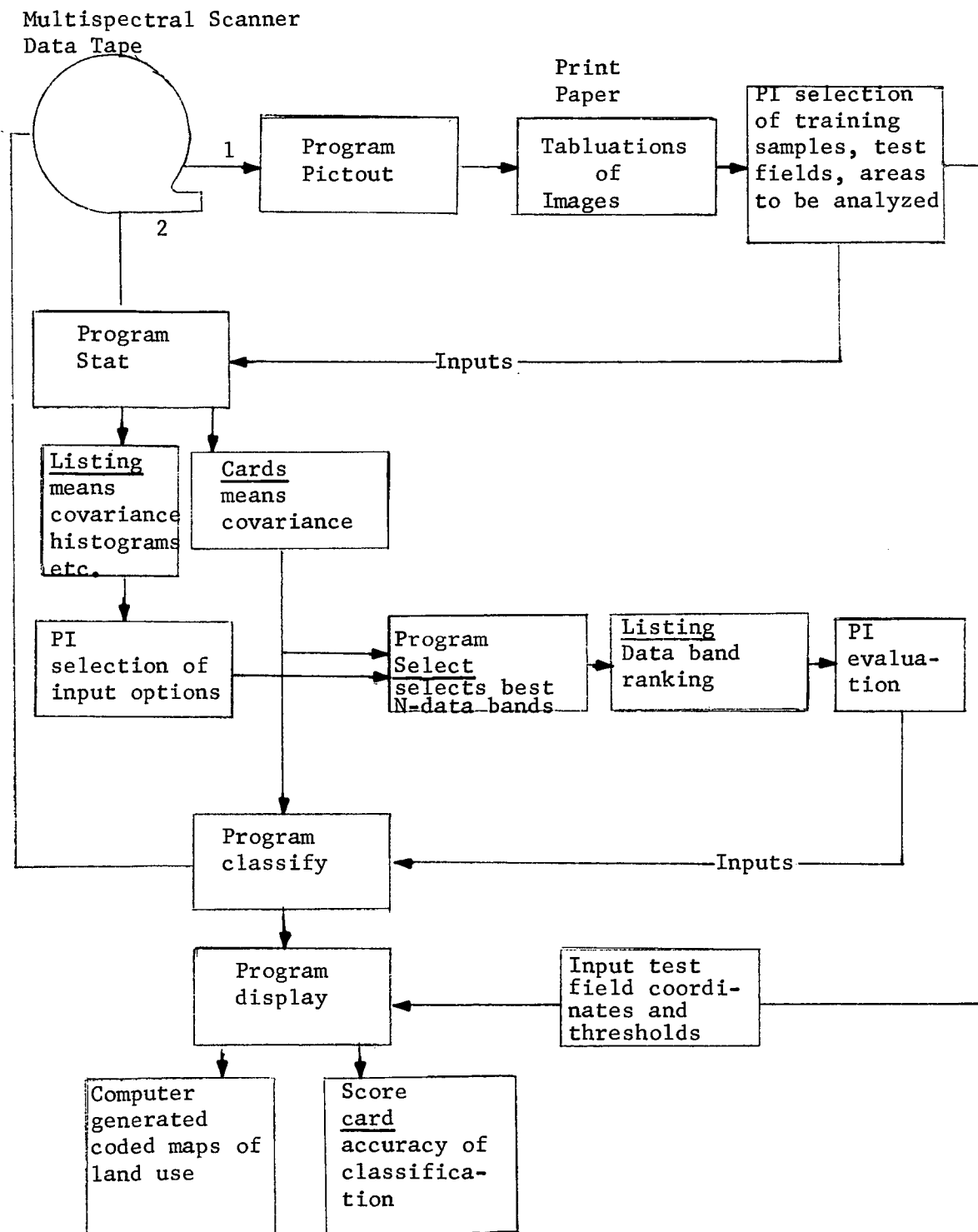
Regardless of how well we do in the development of either a pure digital pattern recognition analysis system or a pure analog pattern recognition analysis technique, each will have inherent disadvantages that are not solvable in pure analog or pure digital. By clever programming, the processing speed can be improved in the digital programs with larger and faster computers, but for all digital techniques processing speed will remain prohibitive. The setup time can be improved in the analog system with state-of-the-art systems, but it too will remain prohibitive. A hybrid computer implementation of the analog and digital solutions developed at the University of Michigan and Purdue can avoid both the long analog setup time and the long computation time by accomplishing the setup tasks with the digital portion of the hybrid computer, and by accomplishing the classification tasks with the analog portion of the hybrid computer. Both the University of Michigan and MSC are doing some development work using hybrid computers. I believe that we can expect some great improvements in analysis time and in reduced cost of analysis. I would like to urge that this area of development be continued.

#### SCHEDULE AND MILESTONES

Figure 17 is a schedule and some milestones for pattern recognition analysis development, the acquisition of equipments that will be used to prepare data for pattern recognition analysis, and some of the missions which will have pattern recognition analysis requirements.



## Flow of Data Through LARS Pattern Recognition System



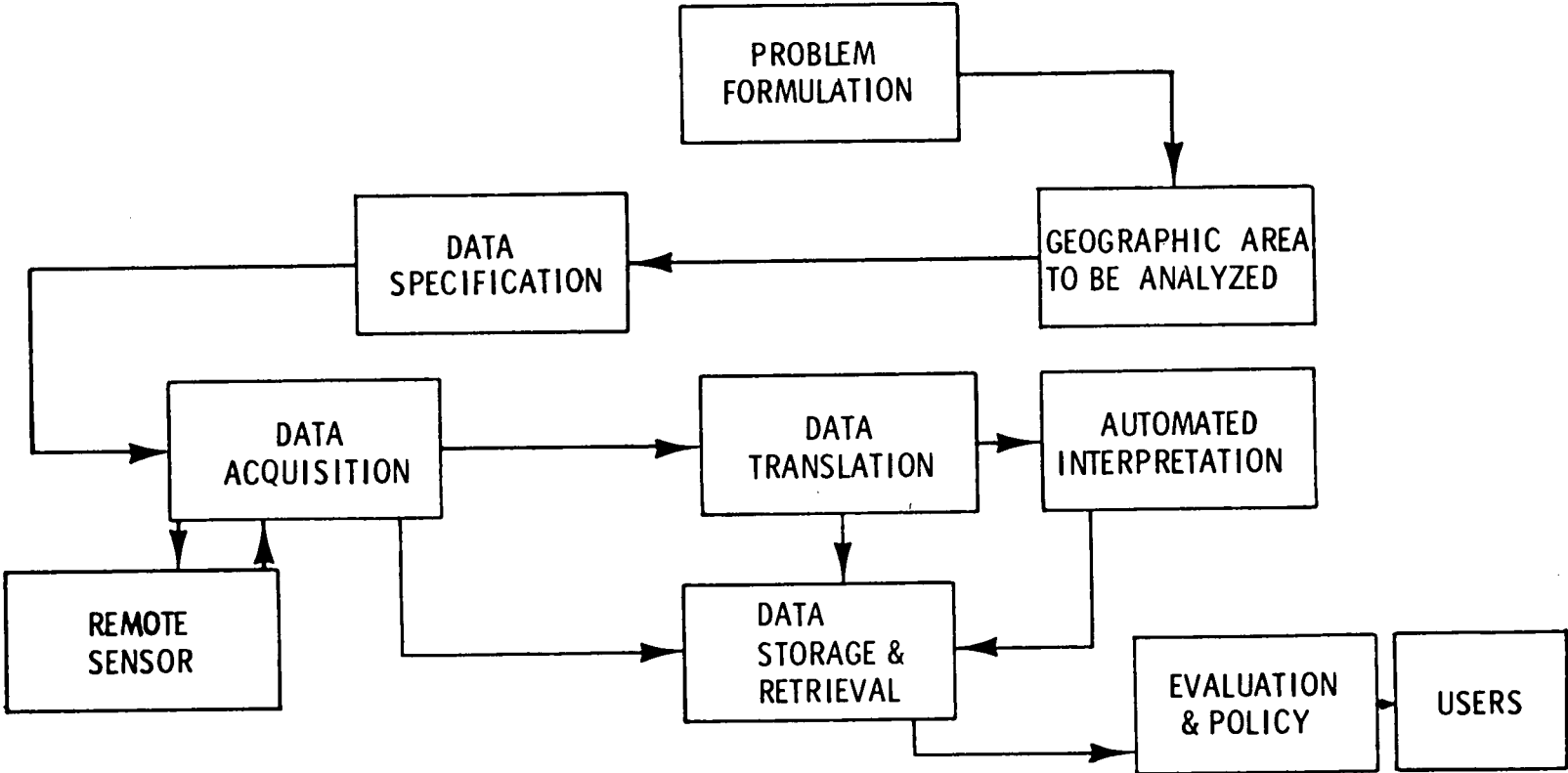
## DETAIL OBJECTIVES

LAND

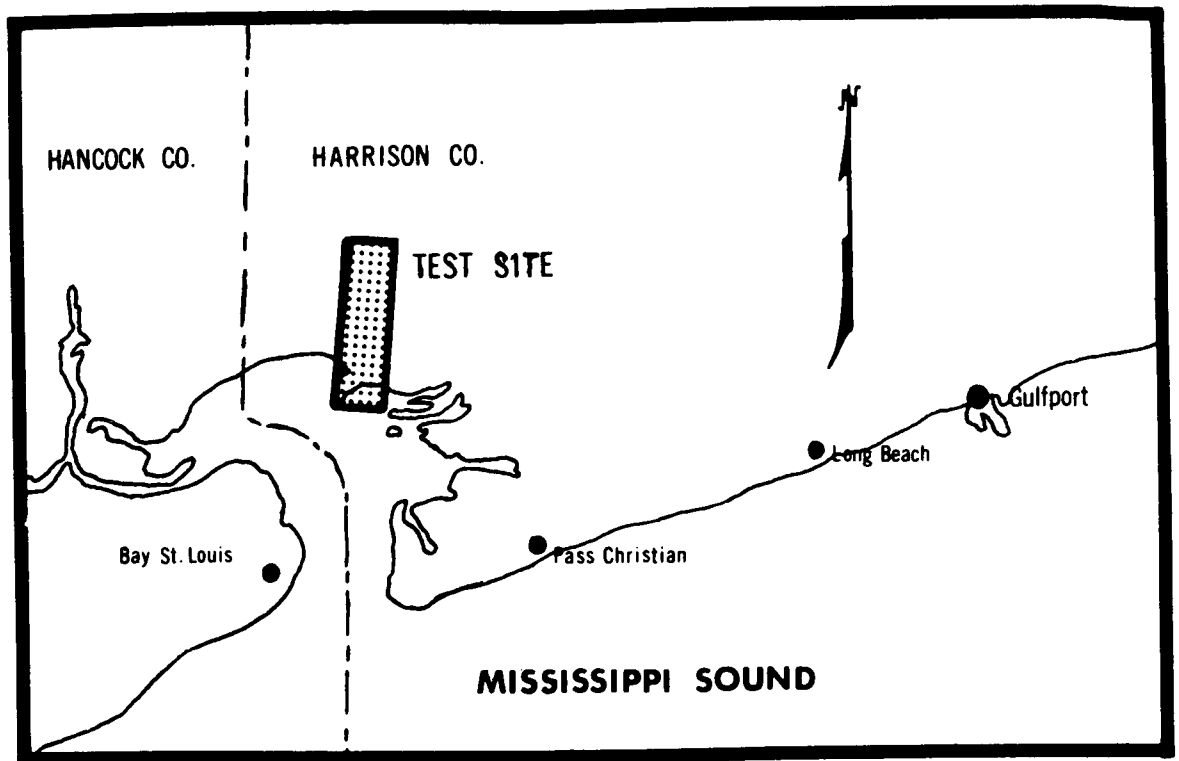
- o EVALUATION OF 24 CHANNEL MULTISPECTRAL SCANNER TO 20K ALTITUDE
- o EVALUATION OF 10 CHANNEL BENDIX SCANNER TO 60K ALTITUDE
- o EVALUATION OF EXISTING AND MODIFIED PATTERN RECOGNITION TECHNIQUES AND SCANNER RESOLUTIONS FOR MIXED TARGET AREAS, e. g. URBAN, FOREST, AGRICULTURE, MARSH
- o DEMONSTRATION OF TECHNIQUE TO APPLICATIONS, e. g. SURFACE CLASSIFICATION FOR LAND USE PLANNING; PLANT COMMUNITY DELINEATION FOR MOSQUITO CONTROL
- o EVALUATION OF "EXTENSION" CAPABILITIES OF TECHNIQUES TO LARGER AREAS OF VARYING GEOGRAPHY
- o DEFINITION OF ELEMENTS OF SYSTEM FOR AREA LAND USE UPDATE SYSTEM, e.g. SATELLITE MEASUREMENTS AND ACCURACIES, COMPUTER OUTPUT FORMATTING

Figure 2

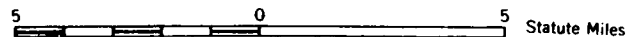
## PROCESSING OF REMOTE SENSING DATA FOR LAND USE INVENTORY



## PILOT AUTOMATED LAND USE UPDATE

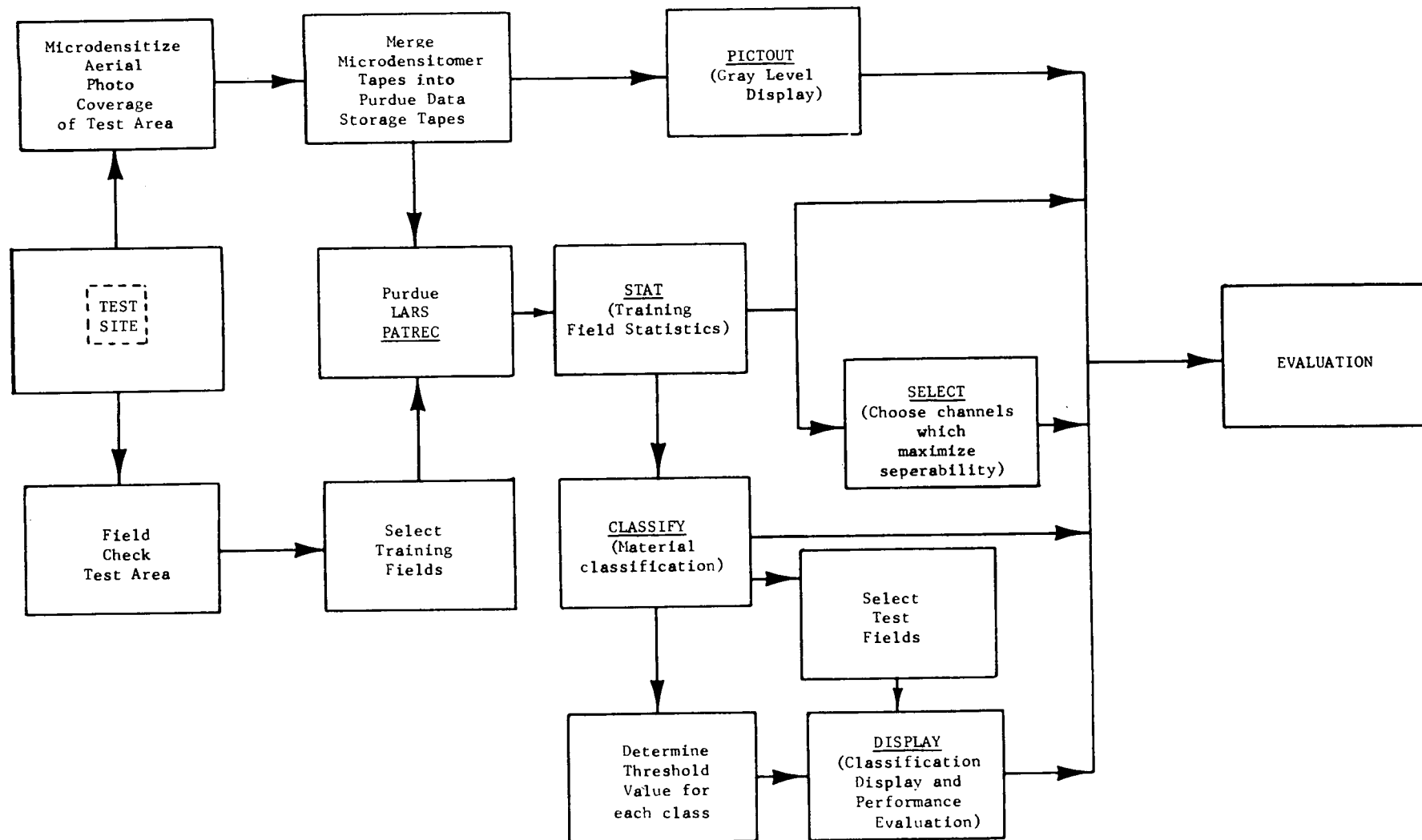


Scale 1:250,000



**Fig.4.** Existing photography covering the 1 x 3½ mile strip of land indicated above was chosen as the test area on which to introduce and exercise automated pattern recognition capabilities at MTF/ERL.

Figure 5 DATA TRANSLATION AND AUTOMATED INTERPRETATION  
SYSTEM EMPLOYED IN PILOT LAND USE DETECTION STUDY



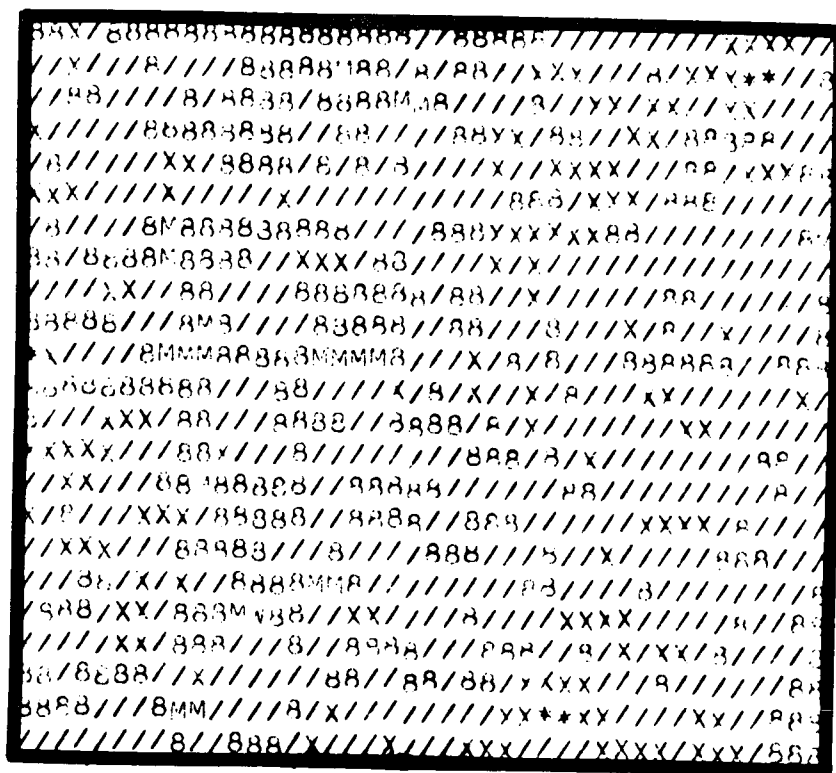


Figure 6. PICTOUT display, channel 2 (.52 - 58 microns).

Symbol	Film Density Interval (relative)
*	89 - 93
X	94 - 96
/	97 - 98
8	99 - 100
M	101 - 103

Pw

RUN NO. 26600061, FIELD 60-1  
 NO OF SAMPLES = 861, FROM LINES 180 TO 200  
 (EVERY 1 LINE(S)), SAMPLES 3 TO 43 (EVERY 1 SAMPLE(S))

THE COVARIANCE AND MEAN FOR TRAINING FIELD 60-1

	.43-	.52-	.62-
	.49	.58	.66
MEAN	100.70	77.79	51.35
ST DEV	1.97	2.14	3.56

COVARIANCE MATRIX

.43 -	.52 -	.62 -
.49	.58	.66
3.90		
2.00	4.56	
.45	.59	12.70

CORRELATION MATRIX

.43 -	.52 -	.62 -
.49	.58	.66
1.00		
.06	1.00	
.06	.08	1.00

06

Figure 7. Example of STAT printout for training field 60-1.

CHANNEL 2            .52 - .58 MICRONS

EACH \* REPRESENTS 16 POINT(S).

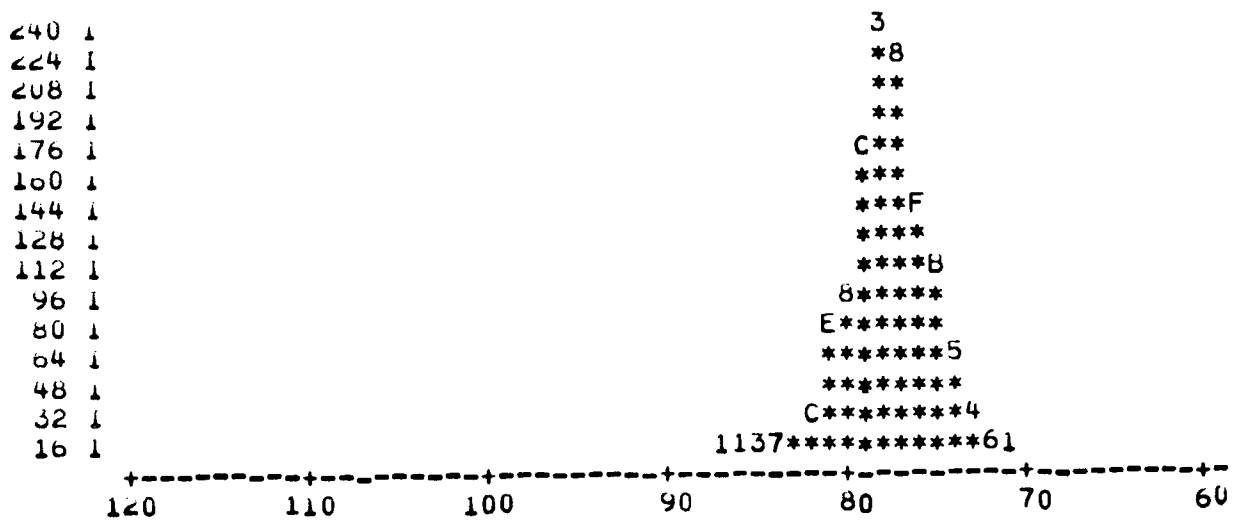
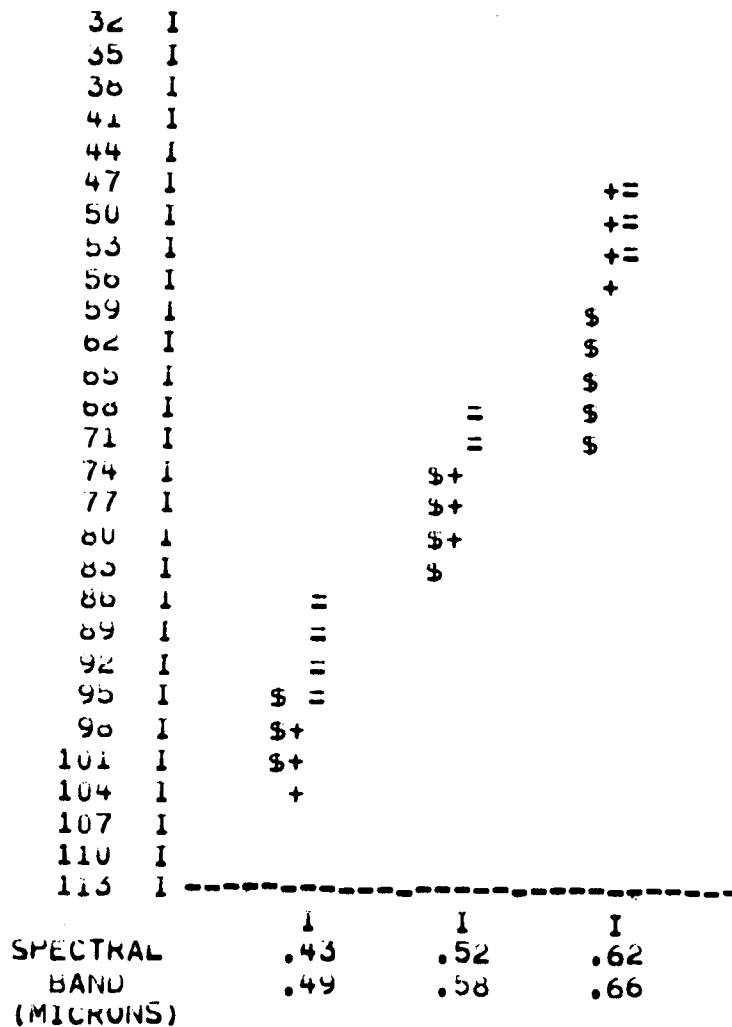


Figure 8 Reflectance histogram for pine (pulpwood), channel 2  
 (.52 - .58 microns).

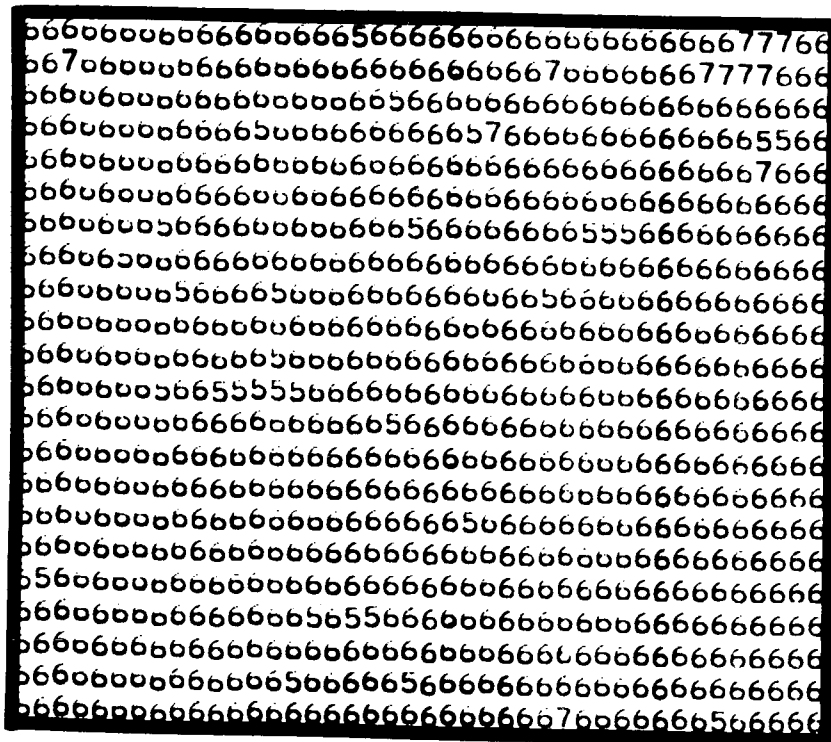




#### LEGEND

- \$ = CLASS HDWOOD
- +
- = CLASS PW
- = CLASS REPRO

Figure 9. Comparison of spectral reflectance of training areas of three classes of material. Reflectance or radiance, increasing upward, is shown for each of three channels used in the pilot study. A vertical line two standard deviations long, centered about the mean radiance, is drawn using alphanumeric symbols.



**Figure 10** Example of CLASSIFY printout for training field 60-1. Using maximum-likelihood scheme actual classification is made for each ground resolution cell.

Symbol	Class
5	Hardwood
6	Pine (pulpwood)
7	Pine (reproduction)

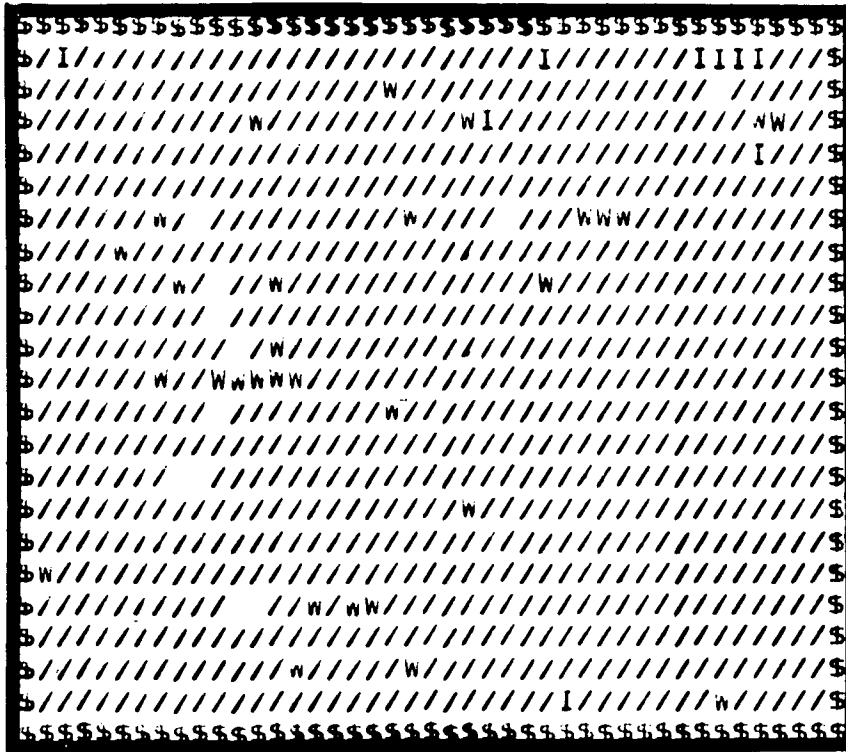


Figure 11 Example of DISPLAY printout of training field 60-1, utilizing the threshold option, the classification made for each ground resolution cell is displayed, and left blank where value at that cell did not exceed threshold value.

Symbol	Classification
W	Hardwood
/	Pine (pulpwood)
I	Pine (reproduction)
Blank	Not classified

# CLASSIFICATION SUMMARY BY TEST CLASSES

	CLASS	NO OF SAMPS	PCT. CORCT	NO OF SAMPLES CLASSIFIED INTO							
				SWAT	MARS	ROW	PAST	HDWO	PW	REPR	THRS
1	SWAT	12113	95.8	11599	0	0	0	0	0	0	514
2	MARS	9968	84.5	0	8424	0	0	756	4	1	783
3	ROW	2141	87.9	0	0	1882	63	0	2	23	171
4	PAST	3050	86.9	0	5	22	2651	6	0	118	248
5	HDWO	1587	89.3	0	90	0	0	1417	71	5	4
6	PW	3566	78.8	0	1	0	0	262	2810	424	69
7	REPR	1656	77.7	0	1	0	90	117	32	1287	129
	TOTAL	34081		11599	8521	1904	2804	2558	2919	1858	1918

OVERALL PERFORMANCE = 88.2

PERFORMANCE BY CLASS = 85.8

Figure 12 Example of a Summarization of Classification Results, Using Training Samples. All fields of a particular material are grouped together and correct recognition percentage is shown. Also given is the number of ground resolution elements being classified into each category.

# DIGITAL TABLE LOOK-UP DATA PROCESSING FLOW

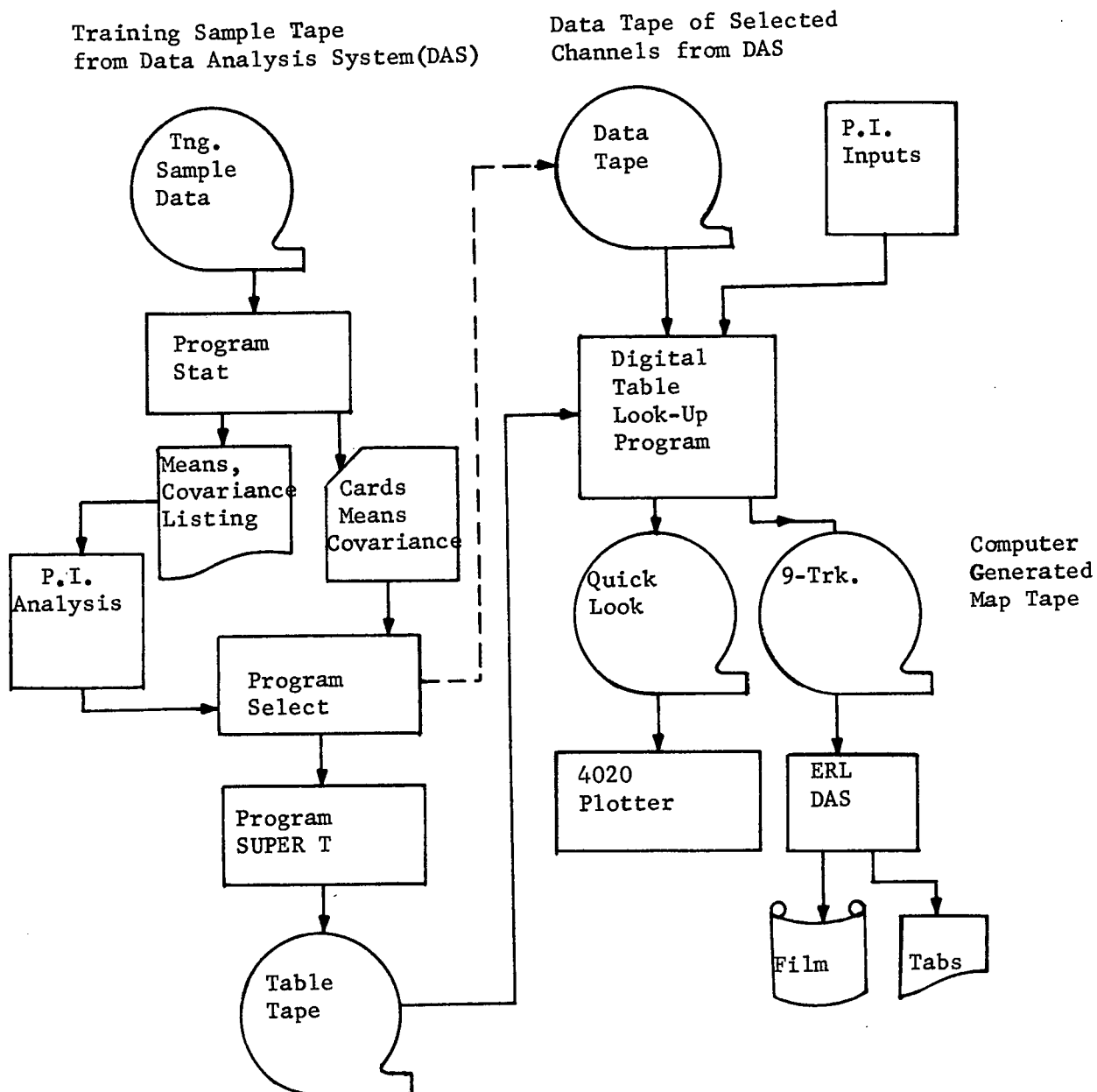


Figure 13

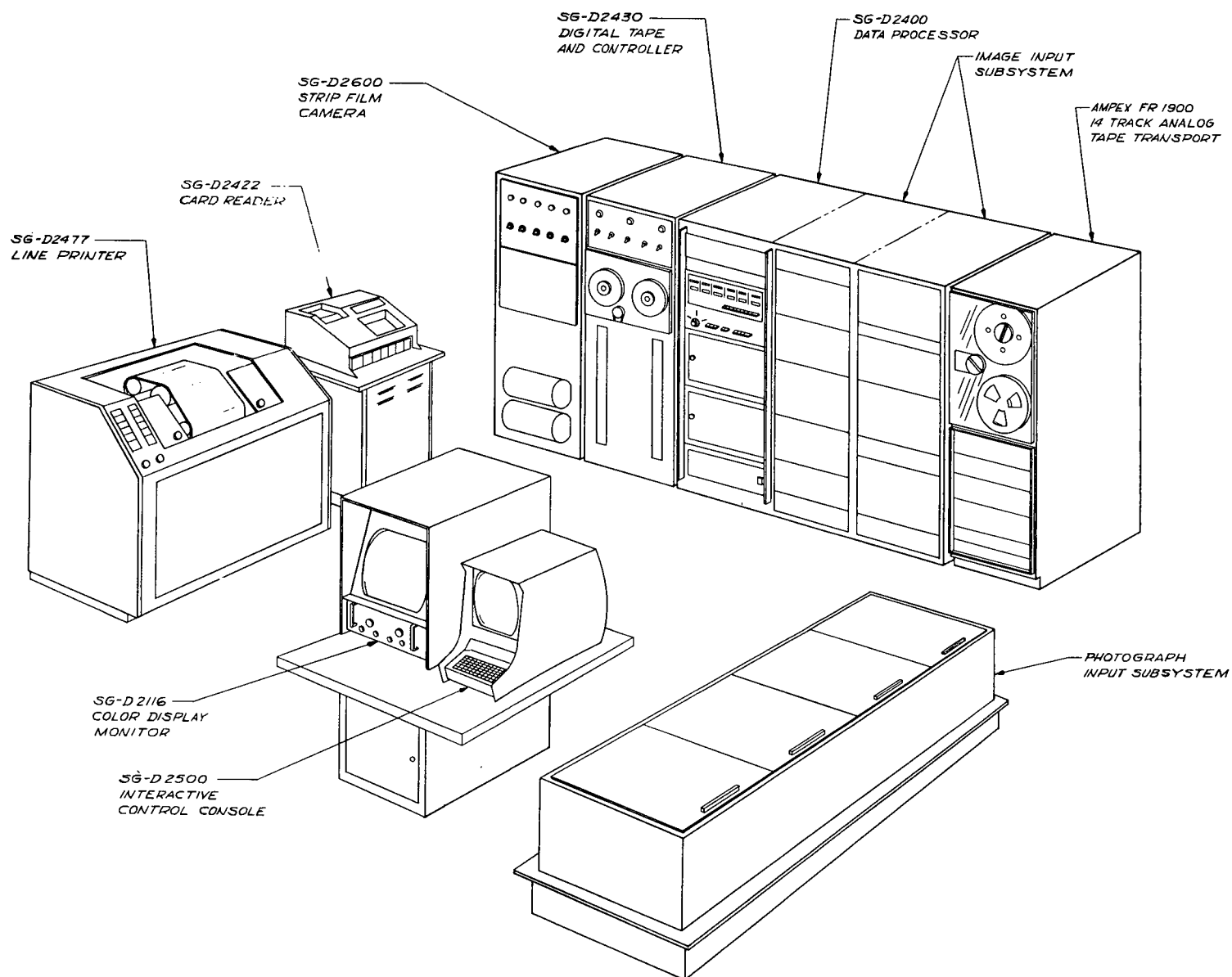


Figure 14 Perspective View of Earth Resources Data Analysis System

# ERL - DATA ANALYSIS SYSTEM

## DATA INPUTS

## DATA PROCESSOR

## OUTPUTS

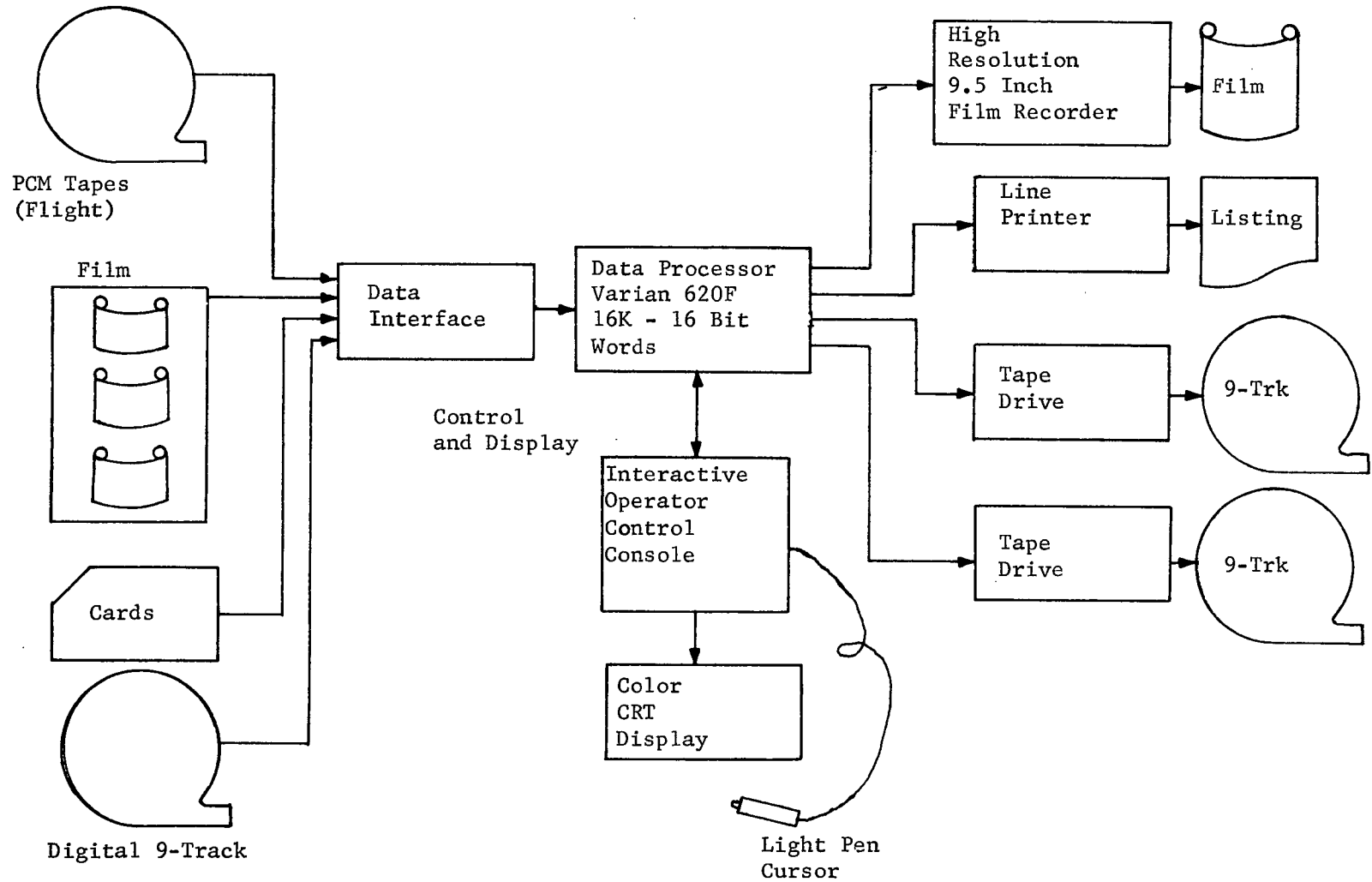


Figure 15

Figure 16

## ERL DAS CHARACTERISTICS

PURPOSE: TO SCREEN, EVALUATE, REFORMAT AND RECORD DATA FROM MULTISPECTRAL SCANNERS, MICROWAVE IMAGERS, MULTIBAND FILMS, AND SCANNING IMAGING SPECTRORADIOMETERS FOR FURTHER ANALYSIS ON LARGE COMPUTERS.

INPUTS: A) FROM TELEMETRY 1-INCH TAPES RECORDED IN BI-PHASE L PULSE CODE MODULATED (PCM) FORMAT  
B) UP TO 4 BANDS OF MULTIBAND CAMERA FILM.  
C) COLOR OR COLOR IR PHOTOGRAPHY FOR TRI-COLOR SEPARATIONS INTO THREE DATA BANDS  
D) IBM CARDS  
E) DIGITAL 9-TRACK TAPES OF IMAGERY

DISPLAY: A) COLOR CRT (525 TV LINE RESOLUTION)  
B) FALLING RASTER MOVEMENT  
C) REFRESHED AT STANDARD TV RATES

CONTROL: INTERACTIVE OPERATORS CONSOLE  
A) SELECTION OF DATA BANDS  
B) LINEAR COMBINATION OF DATA BANDS  
C) REAL TIME COLOR DISPLAY MODIFICATION

DATA PROCESSOR: A) VARIAN 620F COMPUTER  
1) 16K, 16 BIT WORDS  
2) 750 NANO-SECOND CYCLE TIME

OUTPUTS: A) FILM RECORDER  
1) COLOR OR B&W  
2) 9.5 INCH FILM  
3) CONTINUOUS STRIP RECORDING  
B) LINE PRINTER  
C) TAPE DRIVES  
1. 9-TRACK 800 BPI  
2. SPEED 150 IPS



## SCHEDULE AND MILESTONES

CY 72

CY 73

	1st QR	2nd QR	3rd QR	4th QR	1st QR	2nd QR	3rd QR	4th QR	1st QR	2nd QR	3rd QR	4th QR
PATTERN RECOGNITION PROGRAMS		▲										
RECEIPT FROM MSC		▲										
FIRST RUN THROUGH ERL COMPUTERS			▲									
TABLE LOOK-UP DEFINITION AT ERL				▲								
TABLE LOOK-UP OPERATIONAL AT ERL						△						
MODIFICATIONS COMPLETE FOR ERTS							△					
MODIFICATIONS COMPLETE FOR EREP S192										△		
ERL DATA ANALYSIS SYSTEM												
AWARD OR CONTRACT		▲										
PRELIMINARY DESIGN REVIEW			▲									
CRITICAL DESIGN REVIEW				▲								
DELIVERY TO MTF						△						
OPERATIONAL AT MTF							△					
MISSIONS WITH PATTERN RECOGNITION REQ'T's												
MSDS CHECKOUT MISSIONS												
MTF TEST SITE			▲	▲		△						
ATCHAFALAYA TEST SITE				▲		△						
MODULAR MULTISPECTRAL SCANNER EVAL.					△							
ERTS-A							△	△	△	△	△	△
SKYLAB EREP										△	△	△

15-25

TRAINING FIELD PROCESSINGAssumptions:

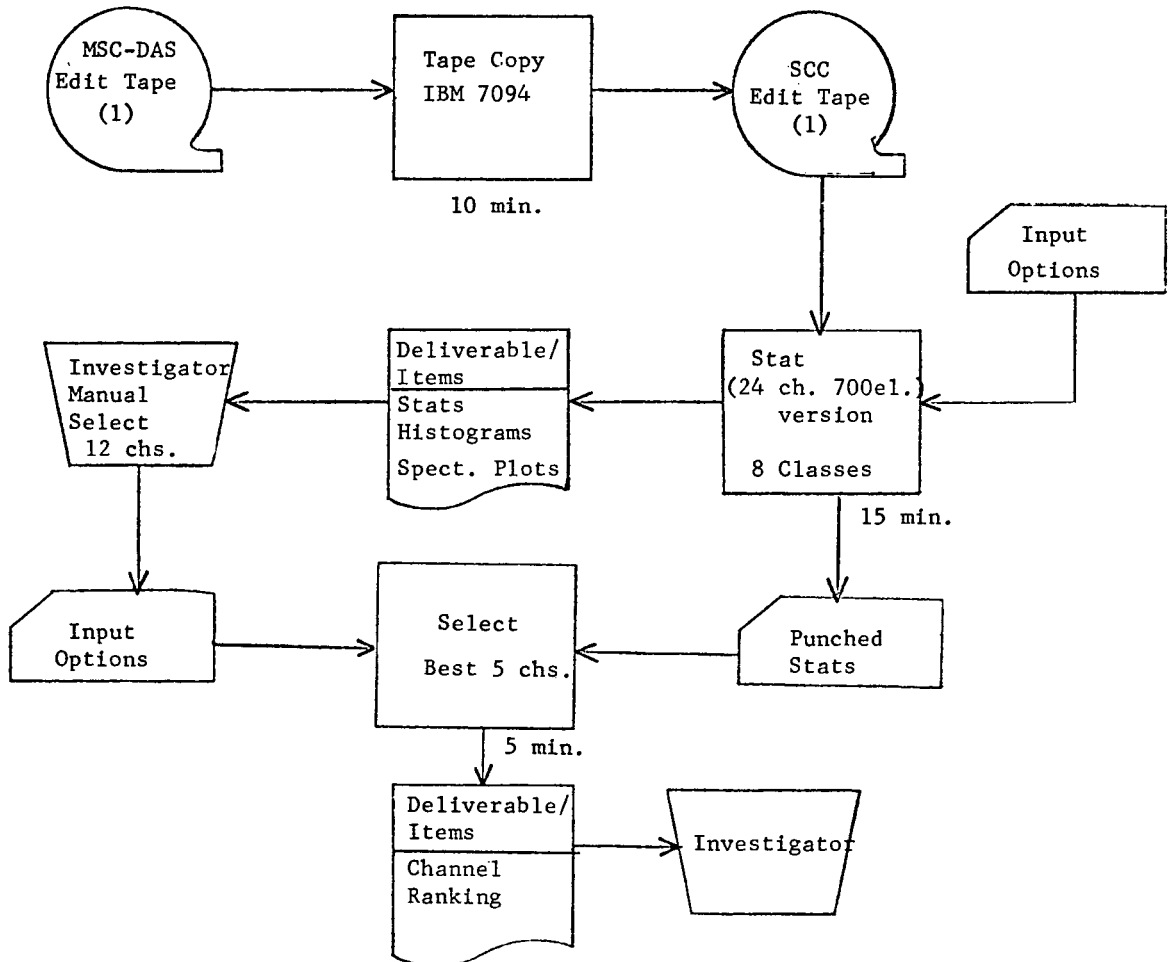
- o Jne 7 track tape contains all training field data.
- o All tapes are error free and properly formatted.
- o Overnight turnaround provided by SCC.

Investigator Participation:

- o Provides histogram and spectral plot scales.
- o Determine best 12 channels.

Computer Processes:

Computer Time	25 min.
Manhours	60 hrs.
Calender Time	1 wk.



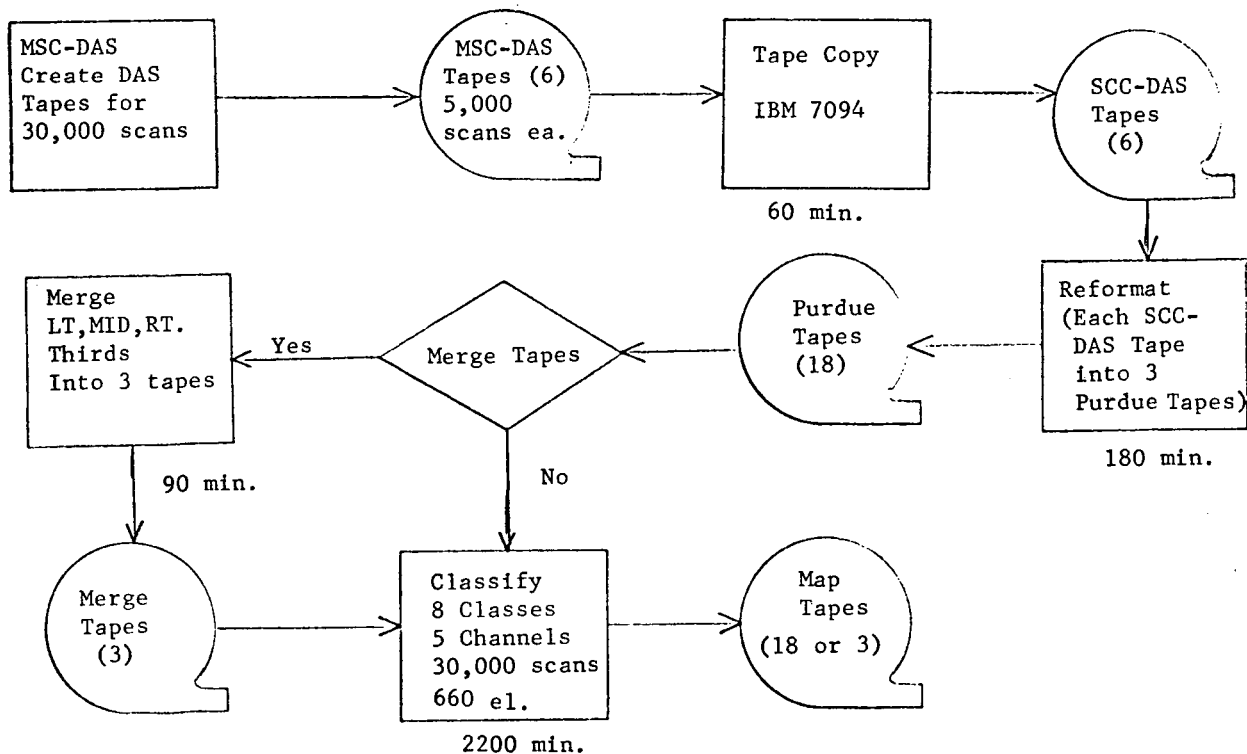
Assumptions:

- 7 track tapes of 5 desired channels.
- All tapes are error free and properly formatted.
- All scans and elements will be classified.
- Overnight turnaround provided by SCC.

Investigator Participation:

- The channels to be used (subset of STAT)
- The classes to be used (subset of STAT)

Computer time	2530 min.
Manhours	200 hrs.
Calendar Time	4 wks.



DISPLAYAssumptions:

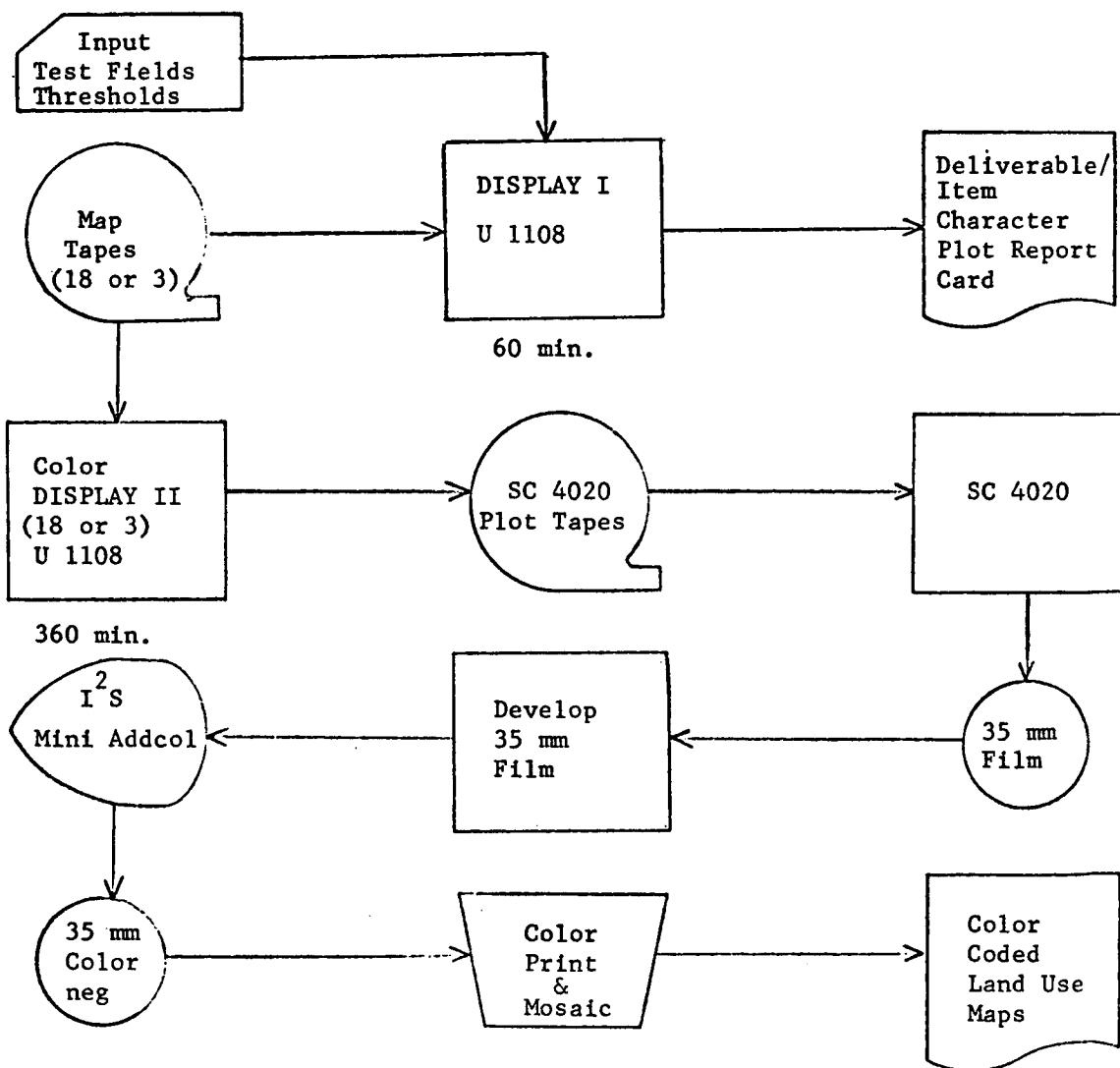
None

Investigator Participation:

- Provide the threshold values for each class.
- Provide test field coordinates.

* Computer time	440 min.
Manhours	370 hrs.
Calendar time	6 wks.

\* Character plot from printer available in one week; 60 min. computer time and 30 manhours.



The time required to complete all processing steps would be :

15-29

<u>Step</u>	<u>Comp Time Min.</u>	<u>Man Hrs.</u>	<u>Cal. Time Wks.</u>
Training Field Processing	25	60	1
Classification	2530	200	4
Display I	60	30	1
Display II	<u>360</u>	<u>340</u>	<u>6</u>
Total	2975	630	12
	49 hrs.35min.	16 wks.	12 wks.

A more detailed description of the steps required to process this data are:

1. Receipt of MSC-DAS edit tapes.
2. STAT (24 ch., 700 ele. version).
3. PI manual select of best 12 chs.
4. SELECT (best 5 of 12 chs).
5. MSC-DAS data tapes for best 5 chs.
6. REFORMAT
7. MERGE
8. CLASSIFY
9. DISPLAY I (character plot)
10. DISPLAY II (color coded land use map)

The flow chart estimates are for a typical mission consisting of 30,000 scan lines of MSS data. It is assumed that eight materials or classes and the best five data channels are to be used.

Basic assumptions are:

1. MSC-DAS Tapes
  - a. Only 7-track tapes are received.
  - b. All tapes are error-free..
  - c. All required data are properly formatted.
2. STAT
  - a. The operational 24 ch., 700 ele. version will be utilized.
  - b. Training field data will be on a single MSC-DAS edit tape.
  - c. STAT will be run for only the training fields contained on the MSC-DAS edit tape.

- d. The PI will provide the following input options:
  - o Histogram high end
  - o Histogram low end
  - o Histogram height
  - o Histogram number of bins
  - o Spectral plot low end
- e. Deliverable items, training classes and fields (for all good chs.) will include:
  - o Means
  - o Covariance matrices
  - o Histograms
  - o Spectral plots
  - o Coincident spectral plots (classes only)
  - o Punched means and covariance matrices on cards (classes only)

### 3. PI Manual Select

The PI will either perform a manual select to determine the best 5 channels or determine which channels (max. of 12) should be utilized by SELECT to determine the best 5 channels.

### 4. SELECT

- a. The production 12 channel LARS Purdue version of SELECT will be utilized.
- b. The PI will provide:
  - o Which training classes and STATS will be utilized.
  - o The number of channel combinations to consider.
  - o Which channels (max. of 12) will be considered.

- o Any weights to be assigned a class.
  - o The number of combinations desired to print the results for.
- c. If statistics are not available for the selected classes and channels (max. 12) in the STAT module training deck, STAT will be required to be re-run for only those selections.
- d. Deliverable item is a listing of the required number of the best channel combinations.

## 5. MSC-DAS Data Tapes

- a. The PI after analyzing SELECTS output will notify the MSC-DAS to generate data tapes for the desired best channels.
- b. No multi-reel tapes.
- c. Only 7-track tapes to be delivered to ERL.
- d. All tapes are error free.
- e. Data is properly formatted on tapes.

## 6. REFORMAT

- a. In order to use the 12 channel, 222 element production version of CLASSIFY, the MSC-DAS data tapes must be reformatted to the Purdue Bulk Data Tape Storage format (i.e., maximum of 12 channels and 222 elements).
- b. Since the production version of CLASSIFY can handle only 110 elements at a time, the REFORMAT tapes will contain data for only 220 elements.
- c. Each MSC-DAS data tape will be reformatted into 3 REFORMAT tapes. Each tape will contain data for 220 elements (approximately 1/3 scan consisting of 700 elements) and be designated LEFT, MIDDLE, and RIGHT thirds).

15-32 7. MERGE

The running of MERGE is optional and beneficial in that it merges the REFORMAT tapes into 3 tapes (one each for the left, middle, and right thirds).

8. CLASSIFY

- a. The current operational 12 channel, 222 element LARS version of CLASSIFY will be utilized.
- b. The PI will provide:
  - o The channels to be used which are a subset of the STAT data.
  - o The classes to be used (subset of STAT data).
- c. All scans and elements will be classified.
- d. Deliverable items
  - o Listing (character plot) of classified data.

9. DISPLAY I

- a. The current operational LARS version of DISPLAY will be utilized.
- b. The PI will provide:
  - o The threshold values for each class.
  - o The desired test fields.
- c. Deliverable items include:
  - o A "report card" giving performance results on a per field and per class basis.
  - o A line printer character plot.

10. DISPLAY II

- a. This is optional and results in a color coded land use map.



- b. The STAT histograms will be utilized to determine bins to be utilized for each color (class).
- c. Combining of colors will be done on the Mini Addcol viewer.
- d. A 30,000 scan line mission will produce 360 frames of color photography. Prints and a mosaic of these photos will be required to produce a color coded land use map.